PRODUCT DEMAND PREDICTION WITH MACHINE LEARNING [APPLIED DATA SCIENCE]

PHASE III PROJECT: LOADING AND PREPROCESSING

TABLE OF CONTENTS:

1) INTRODUCTION
2) PROBLEM DEFINITION
3) DATA PREPROCESSING
    a) DATA COLLECTION
    b) IMPORTNG LIBRARIES
    c) IMPORTING DATASETS
    d) HANDLING MISSING DATA
    e) ENCODING CATEGORICAL DATA
    f) SPLITTING THE DATASET
    g) FEATURE SCALING
4) CONCLUSION
5) REFERENCES

INTODUCTION:

The problem we have taken is prediction of demand for products using machine learning. We know that the demand for products varies with time and place in real life. With demand the product price increases and decreases. We have to create a machine learning model that forecasts product demand based on historical sales data. In this phase' s project we are going to preprocess the data using the historical sales and analyse them.

PROBLEM DEFINITON:

The problem of predicting demand for a new product based on its characteristics and description is critical for various industrial enterprises, wholesale and retail trade and, especially, for modern highly competitive sector of air transportation, since solving this problem will optimize production, management and logistics in order to maximize profits and minimize costs.

DATA PREPROCESSING:

Data preprocessing in Machine Learning refers to the technique of preparing (cleaning and organizing) the raw data to make it suitable for a building and training Machine Learning models.

1. Acquire the dataset
2. Import all the crucial libraries
3. Import the dataset
4. Identifying and handling the missing values
5. Encoding the categorical data
6. Splitting the dataset
7. Feature scaling

DATA COLLECTION:

Dataset required for implementing the project is given below : https://www.kaggle.com/datasets/chakradharmattapalli/product-demand-prediction-with-machine-learning

IMPORTING NECESSARY LIBRARIES:

The dataset contains data about:

1. the product id;
2. store id;
3. total price;
4. base price;
5. Units sold;

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import plotly.express as px
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeRegressor
```

IMPORTING THE DATASETS:

The next step is importing the dataset from the source:

```
data = pd.read_csv('demand_data.csv')
```

HANDLING MISSING DATA:

Lets check if the dataset has any null values and according to that we can delete the entire row or delete it by calculating the mean.

```
data.isnull().sum()
```

Since the dataset has only one missing value, we can use the dropna function to remove the missing value in the column ” total price” by deleting that specific row.

```
data = data.dropna()
```

ENCODING CATEGORICAL DATA:

Categorical data is data which has some categories. We have two categorical data:

Product id;

Store id;

```
label_encoder_x= LabelEncoder()

x[:, 0]= label_encoder_x.fit_transform(x[:, 0])

onehot_encoder= OneHotEncoder(categorical_features= [0])    x=
 onehot_encoder.fit_transform(x).toarray()

labelencoder_y= LabelEncoder()

y= labelencoder_y.fit_transform(y)
```

SPLITTING THE DATASET INTO TRAINING AND TESTING SET:

In machine learning data preprocessing, we divide our dataset into a training set and test set. This is one of the crucial steps of data preprocessing as by doing this, we can enhance the performance of our machine learning model.

Training Set: A subset of dataset to train the machine learning model, and we already know the output.

Test set: A subset of dataset to test the machine learning model, and by using the test set, model predicts the output.

```
x = data[["Total Price", "Base Price"]]

y = data["Units Sold"]

xtrain, xtest, ytrain, ytest = train_test_split(x, y,
                                                test_size=0.2,
                                                random_state=42)

model.fit(xtrain,ytrain)
```

FEATURE SCALING:

Feature scaling is the final step of data preprocessing in machine learning. It is a technique to standardize the independent variables of the dataset in a specific range. In feature scaling, we put our variables in the same range and in the same scale so that no any variable dominate the other variable.

```
st_x= StandardScaler()

x_train= st_x.fit_transform(x_train)
```

```
x_test= st_x.transform(x_test)
```

CONCLUSION:

Thus we have done preprocessing to our data using the seven steps.

REFERENCES:

https://www.javatpoint.com/data-preprocessing-machine-learning

https://thecleverprogrammer.com/2021/11/22/product-demand-prediction-with-machine-learning/